Healthcare Twitter Analytics

Sentiment Analysis: compare two approaches

```
# see http://jeffreybreen.wordpress.com/2011/07/04/twitter-text-mining-r-slides/

# setup
# =====
source("score.sentiment.R")
hu.liu.pos = scan('positive-words.txt', what='character', comment.char=';')
hu.liu.neg = scan('negative-words.txt', what='character', comment.char=';')
pos.words = c(hu.liu.pos, 'upgrade')
neg.words = c(hu.liu.neg, 'wtf', 'wait', 'waiting', 'epicfail', 'mediocrity')

# examples
# ========
sample = c("You're awesome and I love you",
           "I hate and hate and hate. So angry. Die!",
           "Impressed and amazed: you are peerless in your achievement of unparalleled mediocri
ty.")
result = score.sentiment(sample, pos.words, neg.words)
result$score
```

```
## [1]  2 -5  4
```

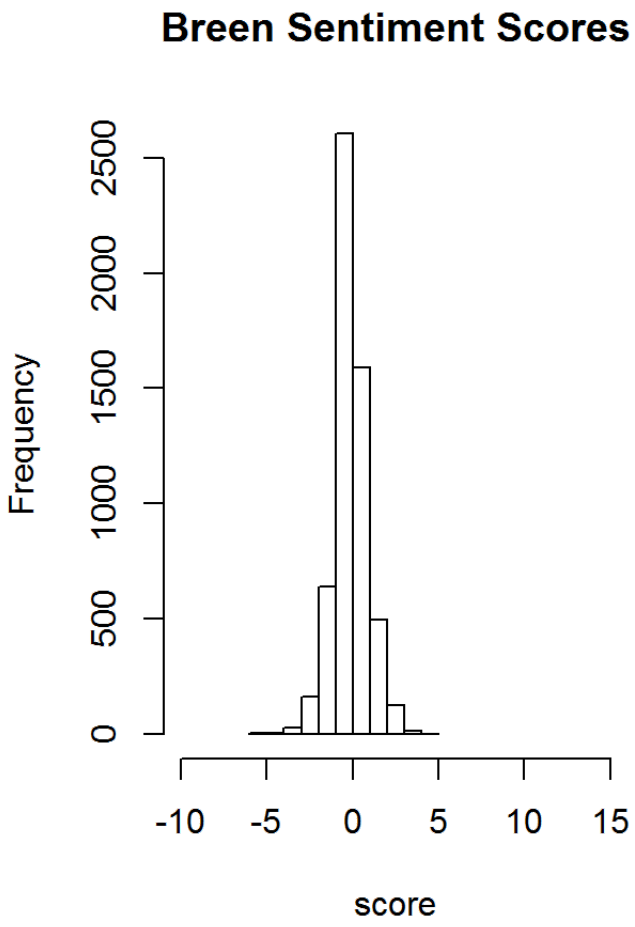
```
score.sentiment(c("@Delta I'm going to need you to get it together. Delay on tarmac, delayed co
nnection, crazy gate changes... #annoyed",
                  "Surprised and happy that @Delta helped me avoid the 3.5 hr layover I was sche
duled for. Patient and helpful agents. #remarkable"),
pos.words, neg.words)$score
```

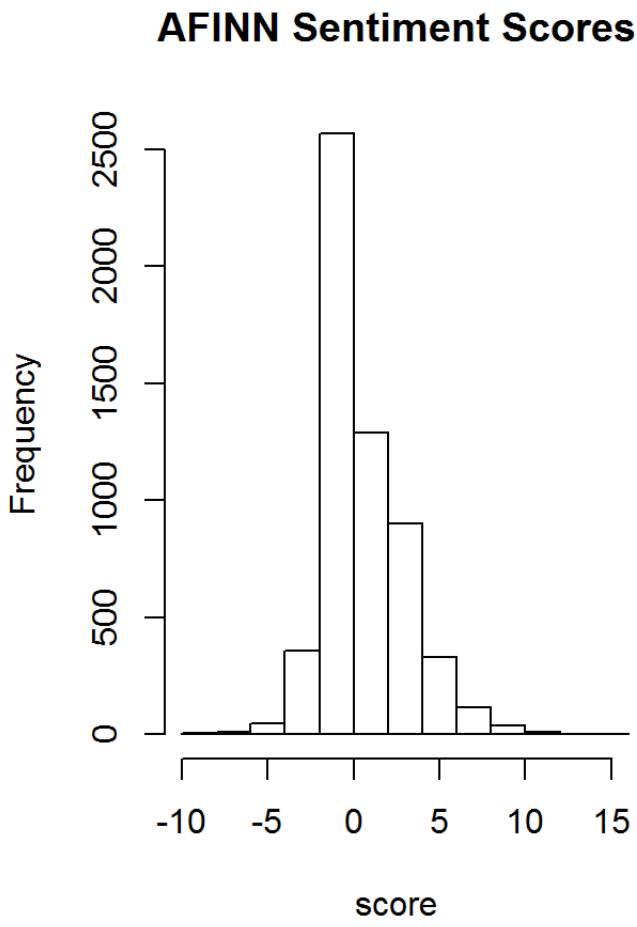
```
## [1] -4  5
```

```
# Look at some tweets
# ==================
data = read.csv("Tweets_Celiac_sent.csv")
text = data$content

# create the sentiment score with the Jeffrey Breen method
scores = score.sentiment(text, pos.words, neg.words, .progress='none')

# append the AFINN sentiment scores
scores$AFINN = data$sentiment
rm(data,text)

grfsummary(scores)
```

```
##            score   AFINN
## n         5675.00 5675.00
## nNA          0.00    0.00
## min         -6.00   -9.00
## mean         0.34    1.21
## median       0.00    0.00
## stdev        1.04    2.34
## skew        -0.02    0.86
## npskew       0.33    0.52
## kurtosis     1.96    2.05
## max          5.00   16.00
```

```
# plot the two sentiment distributions
par(mfrow=c(1,2))
hist(scores$score,main='Breen Sentiment Scores',xlab='score',ylab='Frequency',xlim=c(-10,15))
hist(scores$AFINN,main='AFINN Sentiment Scores',xlab='score',ylab='Frequency',xlim=c(-10,15))
```


Breen Sentiment Scores
Frequency
0
500
1000
1500
2000
2500
-10
-5
0
5
10
15
score


AFINN Sentiment Scores
Frequency
0
500
1000
1500
2000
2500
-10
-5
0
5
10
15
score